IV. ISSUES CONCERNING THE DEVELOPMENT AND USE OF AN ASSESSMENT METHODOLOGY

A. What constitutes an assessment methodology?

The assessment methodology constitutes the decision process (including principles of science, statistics and logic used in interpreting data and information relevant to segment conditions) that a state employs to determine to which of the five integrated reporting categories a segment belongs. It is important that assessment methodologies must be consistent with applicable WQSs. They should also be consistent with sound science and statistics.

As described in section 130.7 (b)(6) (i- iv), each state shall provide documentation to the Regional Administrator at the time that the integrated report or the separate section 303(d) list is submitted. This documentation must support the state's determination to list or not list its segments as required in 130.7(b)(1) and 130.7(b)(2). A major component of this documentation is a description of the methodology that the state used to develop their Integrated Report or the separate section 303(d) list.

The methodology should: 1) explain how the state identifies, considers (evaluates) all existing and readily available data and information; 2) articulate the basics of the quality assurance and quality control (QA/QC) criteria used to evaluate data submitted by outside entities to determine what weight, if any, should be assigned to said data and information; and 3) explain the analytical approaches, including statistical analyses, used to infer true segment conditions from all valid existing and readily available information. The decision processes the states describe in the methodology should provide all stakeholders with the opportunity to understand exactly how assessment decisions are made.

Prior to submission of its Integrated Report, each state should provide the public with the opportunity to review and comment on the methodology, consistent with their continuing planning process (CPP), public participation policies, and monitoring strategies.

B. What will EPA do with the methodology?

When a state has by rulemaking adopted a methodology as part of its approved water quality standards and the water quality standards are applicable for CWA purposes, 40 CFR § 131.21, EPA will apply the approved methodology as it reviews the state's submission in order to determine whether to approve or disapprove the section 303(d) list (Category 5). If a state has not by rulemaking adopted a methodology into its water quality standards, EPA will consider the state's methodology, to the extent that it reflects a reasonable interpretation of the state's water quality standards and sound science, in determining whether to approve or disapprove the section 303(d) list. In either scenario, EPA encourages the state to make available the most recent methodology used to develop the current Draft 2006 Integrated Report (or separate 303(d) list) prior to submission of their IR. The methodology allows EPA and other reviewers to understand the decision process followed by the states as they review the 2006 Draft. Where EPA has concerns with the assessment methodology, EPA will provide comments to the state to assist in developing an approvable section 303(d) list.

For methodologies that are not part of the state's applicable water quality standards, EPA will consider the methodology as it assesses whether the state conducted an adequate review of all existing and readily available water quality-related information, whether the factors that were used to make listing

and removal decisions were reasonable, whether the process for evaluating different kinds of water-quality related data and information is sufficient, and whether the process for resolving jurisdictional disagreements is sufficient. If EPA finds that the state's methodology is inconsistent with its water quality standards, and its application has resulted in an improper section 303(d) list, EPA may disapprove the list. Regardless of the suitability of the methodology, EPA must review the list for consistency with the relevant provisions of the CWA and the regulations.

EPA sees the methodology as an evolving document which states periodically revise as appropriate at some time during the listing cycle[11]. As such, EPA strongly encourages states to submit their draft and current methodologies to EPA and to the public for review and comment (but not formal approval) well in advance of any deadline the state sets for submission of data and information.

C. Data Assembly

40 CFR section 130.7(b)(5) requires that "Each State shall assemble and evaluate all existing and readily available water quality related data and information to develop the list."

States should solicit data and information including, but not limited to, the types listed below:

- observed effects (see glossary)
- closures, restrictions and/or advisories applicable to swimming, fish consumption, and drinking water
- violations of Safe Drinking Water Act (SDWA) standards
- segment-specific ambient monitoring-chemical, physical, and/or biological
- large-scale probabilistic monitoring designs
- simple dilution calculations
- predictive (simulation) modeling,
- landscape analysis
- remote sensing
- complaints and comments from the public

To the extent practicable, such types of data and information should be drawn from existing compilations of information regarding water quality, including, but not limited to:

- publicly-available databases (e.g., STORET)
- source water assessments per the Safe Drinking Water Act
- monitoring information from pesticides registrations
- watershed plans and other kinds of water quality or natural resource management plans
- Superfund Records of Decision
- reports prepared pursuant to sections 305(b), 303(d), 314, and 319 of the CWA

[11] EPA notes that it does not view state regulatory provisions concerning assessment methods that merely describe the sufficiency or reliability of information necessary for states to make an attainment decision as water quality standards, because they do not have the effect of changing the ambient conditions (i.e., magnitude, duration, frequency of concentrations of pollutants) considered necessary to support a designated use.

Such types of data and information should also be solicited from a wide variety of organizations and individuals, such as:

- other state agencies such as Fish and Wildlife, Parks, and Agriculture Departments
- federal agencies, including EPA, USGS, NOAA, USDA, and USFWS
- local governments
- drinking water utilities and state agencies responsible for SDWA implementation
- universities and other research institutions
- environmental consulting firms
- National Pollutant Discharge Elimination System (NPDES) permittees
- conservation/environmental organizations
- outdoor recreation organizations
- citizen monitoring groups

EPA regulations provide that states should actively solicit organizations and individuals such as those listed above. See 40 CFR 130.7(b)(5)(iii). EPA considers active solicitation as notifying local, state, and federal agencies, members of the public, and academic institutions that the state is seeking water quality related data and information for the purpose of developing the Integrated Report, through notices in the State Register, notices or announcements in appropriate local or trade papers, direct mailings to members of the public that have previously submitted public comments or other interested parties on the State's mailing list, or announcements and requests for data at appropriate public meetings or informational meetings. EPA recommends that states also request such data and information via letters sent to other state agencies, federal agencies and academic institutions that may have data/information.

If the state has specifications for data and information, these specifications should be included in any requests for information. To facilitate the timely completion of a draft list that can be distributed for public review and comment, states may set a reasonable "cut-off" date after which no additional data or information will be considered in the preparation of the draft section 303(d) list and other aspects of a preliminary Integrated Report. If a state institutes a cutoff date for data submission, effective prior to establishing a draft list, there could also be a separate data solicitation step prior to compilation of a final 303(d) list. Under this scenario, the state would compile the preliminary list using all information it has at hand based on identified data sources. Additional data submissions during the public comment period would then be evaluated, appropriate changes to the draft list would be made based on these new data or information.

If the state intends to consider only data and information submitted prior to a certain cutoff date, the state should clearly explain that this is the only opportunity for the public to provide data and information for the current assessment cycle, and that data submitted after that cutoff date would be considered during the next listing cycle. States should provide a mechanism for an exception to the limit for the submission of data if the submitter can demonstrate that the data were readily available prior to the data cutoff date and should have been included in any reasonably diligent state review of data. EPA will generally limit its review of a state listing submission to the data and information assembled by the state prior to the data cutoff date if the state was reasonably diligent in assembling available data and information and soliciting data and information from the public.

EPA is aware that many states have turned to the rotating basin strategy as a technically sound approach for making assessment determinations of the state's waters. In this approach, the available monitoring resources are concentrated or targeted in one portion of the state for a specified period of time, thus allowing for data to be collected and assessed in a spatially and temporally focused manner. Over time, every portion of the state is targeted for this higher resolution monitoring and assessment effort (often over a five-year period), however the state must consider all existing and readily available data and information during the development of its 2006 Integrated Report, regardless of where in the state the data and information were generated.

The state should make reasonable efforts to obtain and consider sources of data and information not provided by commenters. If particular data/information referenced in the public comments are not provided, EPA expects states to make a reasonable effort to secure the data. Solicitation requests should note that at a minimum commenters should provide as much information as possible in order for the state to be able to obtain the data or information, and again emphasize any state criteria for considering and prioritizing data sets.

D. How should the methodology describe data and information expectations?

1. Data Quality Considerations

A state must evaluate all existing and readily available data and information, to establish how it should be used in attempting to make a WQS attainment status determinations, applying reasonable and scientifically sound data evaluation procedures. Such evaluation protocols should strike a balance between: (1) employing only the very highest quality data, and (2) employing as much useful information about the condition of as many segments as possible. That is, these protocols should reflect both legitimate concerns about basing decisions on the best possible information and the fact that there is relatively little or no segment-specific monitoring data or other forms of assessment-relevant information available for the majority of the nation's waters.

Such protocols/evaluation criteria would include typical elements of a quality assurance project plan (QAPP). Examples of such elements include a description of the methods used to collect the data in the field, a description of the methods to assure proper handling and "chain of custody" of the samples during transport to the laboratory, documentation of the laboratory methods used to perform the analysis of samples, and a description of any independent audits to verify the consistency of the data. In their articulation of QA/QC expectations for data and information submitted by others, states should describe the types and amount of metadata that should be provided along with specific sets of data and information. If an outside entity fails to provide necessary metadata along with submitted data and information, the state should attempt to obtain the metadata from the data-submitting organization before concluding that the data and information is of low quality, simply due to lack of metadata.

Data quality criteria should be published along with any solicitations of data and information. Ideally, such QA/QC protocols should be made available to the public well in advance of any such solicitation for any given IR reporting cycle.

In addition to articulating their data review criteria, EPA recommends that states work with data-generating organizations not only during the period of time just before the Integrated Report development, but on a more continual basis, to help ensure their data are collected and stored in such a way that the data will of high quality. States may wish to encourage such organizations to develop QAPPs and submit them to the state for review and comment, and even perhaps formal approval by the state. A state may elect to employ a rebuttable presumption that data and information submitted by organizations with a state-approved QAPP meets the state's QA/QC standards. Lack of a State-approved QAPP should not, however, be used as the basis for summarily rejecting data and information submitted by such organizations, or assuming it is of low quality, regardless of the actual QA/QC protocols employed during the gathering, storage, and analysis of these data.

2. Data Representativeness Considerations

The spatial and temporal representativeness of data and information should be considered by states as they attempt to characterize conditions in a given segment. Clearly, the degree of confidence in a WQS attainment status determination increases as the amount of data and information grows. Ideally, all decisions about the WQS attainment status of individual assessment units would be based on a complete census of water quality conditions, which could involve sampling every portion of a waterbody at frequent intervals. Unfortunately, gathering this vast amount of data is not currently feasible, due to the limitations of current monitoring technology as well as the amount of funding available for gathering and analysis of water quality information.

Even for those segments where unusually large amounts of monitoring data is available, compared to most waterbodies, the percentage of all possible locations in time and space from which data has been collected is very, very small. Given this situation, states and EPA will continue to need to make WQS attainment status determinations by extrapolating, in time and space, to a substantial degree, from individual points of data.

Hence, state methodologies should describe, in general terms, the decision logic used to determine the temporal and spatial extent a grab sample can be construed to represent. In order to make credible assessment determinations, states should employ approaches that strike a balance between the extremes of: (1) considering every grab sample to be representative of merely the instant in which, and the drop of water from which, each was taken; or, (2) assuming each such sample is representative of conditions over several years, and covering hundreds of stream miles or hundreds of lake acres. (Note that available data and information should be used to assess attainment of applicable water quality standards unless a specific technical rationale is provided to support a determination that such data and information should not be used (see 40 CFR 130.7(b)(6)(iii-iv)).

Many state numeric water quality criteria include multiple day averaging periods, while most state monitoring programs do not collect samples at a rate of one or more per day. In such circumstances, states should decide how far out in time to extrapolate from the time at which a particular single grab was collected. EPA recommends that such decisions be based on contextual information regarding conditions when and where the grab was taken. For example, such information might include: 1) precipitation, 2) streamflow, 3) location of point source discharges in relation to the monitoring site, 4) land use patterns in the vicinity, 5) expected patterns of pollutant loading from the different kinds of sources present in the

watershed, 6) occurrence of a chemical spill or other unusual event, and 7) historic patterns of pollutant concentrations in the monitoring segment and/or waterbodies similar to it.

For instance, such contextual information might indicate that levels of a pollutant under study are likely to have remained fairly constant over a certain period. This would generally be a reasonable conclusion if, for example, available information suggested that both pollutant loadings and stream flow remained fairly steady over that period. In such cases, it could be reasonable to assume that the concentration seen in a sole available grab sample was representative of average conditions over the period of interest. On the other hand, if it were known that the watershed draining into a segment had a large number of precipitation-dependant sources of pollutants, a particular sample had been collected during the only significant rainfall that occurred during that period, and the precipitation event was of a duration shorter than the averaging period used in the water quality standard, then it could make sense to conclude that the concentration in that sample was not roughly equal to the average over the period in question (e.g., 1 day, 4 days, 7 days).

Similarly, contextual information can help inform a decision as to how far out in space to extrapolate from a particular sampling point. Where no point source dischargers are present and land use practices are the same over large areas, assuming data collected at a particular monitoring site is representative of conditions over a long stretch of river could be reasonable. On the other hand, if a number of point sources and a variety of nonpoint sources are found along a similar length of stream, it may be prudent to presume that data from a particular sampling site is representative of only a portion of that river reach.

In deciding how broad a span of time and space to assume a particular grab sample might represent, States may wish to consider the implications of a more expansive versus a more cautious approach to interpreting available monitoring data in the context of available metadata. Willingness to extrapolate further in time and space will generally lead to making WQS attainment/non-attainment determinations on a larger number of waters and designated uses, while a more cautious approach can result in a higher proportion of waters and uses being reported as "status unknown".

Though a determination of whether a single grab sample can reasonably be construed to be representative of (i.e., close in value to) average conditions over a specified period is an important step in the assessment process, the mere fact that the only grab sample available for a particular period is not deemed representative of average conditions over said period does not necessarily mean that it could not be used as the basis of a WQS attainment status determination. For instance, despite being non-*representative* of the average concentration, it may be indicative of the average, or at least a fairly reliable indicator of whether or not the average concentration in the waterbody over said period is above or below the level specified in the WQS.

For example, it is widely known that dissolved oxygen levels rise and fall in most waterbodies following a diurnal cycle. Hence, if a grab sample were collected at 5 a.m. (around when DO levels should be at the lowest point during the daily cycle) and the DO level in the sample was above, or even somewhat below, the level specified in an applicable WQC expressed as a 24 hour average concentration, it would be reasonable to assume the daily average concentration of DO on the day the one grab was collected was higher than that specified by the WQC. (Conversely, if a DO sample were collected at 6p.m., i.e., during the *high* end of the diurnal cycle of DO levels, and the concentration was below, or

even slightly above, the specified concentration, it could be reasonable to conclude that the 24-hour average DO level was below that specified by the WQC.)[12]

Awareness of the types of sources upstream of a site and knowledge of the weather at the time of sampling can also be instructive. For instance, if the level in the sole grab of a pollutant associated primarily with nonpoint sources was slightly higher than the criterion-concentration, but the grab had been taken during a one-in-10-year one-hour rainfall event, it could be reasonable to assume the 4-day average was lower than the criterion-concentration.

Similarly, EPA believes that data should not automatically be treated as unrepresentative of relevant segment conditions solely on the basis of its age without supporting information indicating that the data are not a good indicator of current conditions. However, older data should be evaluated with care. For example if the most recent data for a particular assessment unit is 10 years old, and that data indicated that average and/or peak conditions in a segment at that time were worse than those specified by an applicable WQC; and, since that time, all the sources of the pollutant in question had been required to dramatically lower the levels of the pollutant in their effluent, and few changes that would lead to increased loadings of the pollutant had taken place in the watershed, it could be reasonable to assume that the segment was now meeting the WQC for that pollutant. By contrast, if 15 year old data indicated that a segment was then just barely meeting WQS for several pollutants associated with urban runoff, and the watershed of that segment had since that time undergone considerable urbanization, a conclusion that the segment was no longer meeting WQC for some or all of those pollutants could be warranted.

States should be cautious about employing assessment methodologies that assign little or no weight to data consistent with state QA/QC protocols based on the theory that it is "unrepresentative" simply because the data seem to reflect unusual circumstances. Rather, such unusual circumstances should be evaluated in the context of the specific requirements of applicable WQSs. In assessing potential adverse effects on humans or other life forms, it is just as important to be cognizant of potential short term events as it is to reflect longer term "average" conditions. Short term exposure to very high levels of pollutants (or low level of necessary elements like oxygen) can be extremely harmful, even lethal. For this reason, EPA and state WQC for a number of pollutants include concentration/duration combinations for short periods as well as such combinations for longer periods. Such criteria are typically referred to as acute and chronic WQC, respectively.

Extreme values or "outliers" can be very relevant when dealing with WQC aimed at protecting humans or other life forms against adverse effects of acute (short term) exposure to pollutants. The fact that such values may occur fairly infrequently and are not representative of long term average conditions is unimportant when dealing with WQC expressed as short-term that should occur only rarely, if ever. EPA's WQC addressing acute exposure of freshwater aquatic life to toxic chemicals are an example of WQC expressed in this way – they are one-hour average concentrations that should be surpassed no more than once every three years on average. WQC expressed as instantaneous concentrations never to be surpassed address even more rare, but nevertheless harmful, conditions.

[12] Another example of using data sets consisting of just one grab sample to make inferences about conditions over longer periods is illustrated by one aspect of the 1986 EPA water quality criteria document for bacteria (www.epa.gov/waterscience/beaches/1986crit.pdf).

Caution regarding exclusion of "outliers" is expressed in EPA's *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9)* (EPA/600/R-96/084) published in July 2000, available at http://www.epa.gov/quality/qa_docs.html):

> "One should never discard an outlier based solely on a statistical test. Instead, the decision to discard an outlier should be based on some scientific or quality assurance basis. Discarding an outlier from a data set should be done with extreme caution, particularly for environmental data sets, which often contain legitimate extreme values. If an outlier is discarded from the data set, all statistical analysis of the data should be applied to both the full and truncated data set so that the effect of discarding observations may be assessed. If scientific reasoning does not explain the outlier, it should not be discarded from the data set." (EPA/600/R-96/084, pp. 4-26).

Additional guidance about "outliers" can be found in the discussion of trimmed means on page 35 of *Biological Criteria: Technical Guidance for Survey Design and Statistical Evaluation of Biosurvey Data* (EPA/822/B/97/002).

However, disregarding valid data gathered during extreme conditions (e.g., significant droughts or floods) can be appropriate if applicable state's WQS include a provision specifying that some or all WQC do not apply during certain rare events, such a 7Q10 low (or high) stream flow. Also, data collected at certain times of years could legitimately be disregarded when making use support status determinations based on seasonal WQC – ones that apply only to times of year other than that when these particular data were collected.

In addition to such "temporal waivers" of WQS, state regulations often contain "spatial wavers"– portions of segments in which some or all WQS do not apply. Mixing zones in the immediate vicinity of NPDES-regulated discharges are the most common example of such exemptions. Hence, data collected within the confines of designated mixing zones should not be applied against some or all WQC otherwise applicable to the receiving segment.

3. Data Quantity Considerations

EPA encourages the collection of adequate data to make well-grounded attainment determinations. EPA has not established, required, nor encouraged the establishment of rigid minimum sample set size requirements in the WQS attainment status determination process. EPA is particularly concerned with application of such thresholds state-wide, without regard to key factors like the manner in which applicable WQC are expressed, variability in segment-specific conditions, and fluctuations in rates of pollutant loading. Rather if employed, target sample set sizes should not be applied in an assessment methodology as absolute exclusionary rules, and even the smallest data sets should be evaluated and, in appropriate circumstances, used. While it may be appropriate to identify target sample sizes as a methodology is developed, states should not exclude from further consideration data sets that do so solely because they not meet a target sample size. A methodology may provide for an initial sample size screen, but should also provide for a further assessment of sample sets that do not meet the target sample size. (EPA suggests that states avoid setting target sample set sizes higher than the amount of data available at most sites.)

Assessments based on larger sample sets are more likely to yield accurate conclusions than assessments based on smaller sample sets. For example, smaller sample sets are more prone to lead to erroneously concluding that at a WQC has not been exceeded, because they result in a lower probability of detecting WQSs exceedances that have actually occurred. (EPA, *Consolidated Assessment and Listing Methodology – Toward a Compendium of Best Practices* (CALM) July 2002, pp. 4-9).

Any target sample set size thresholds must be consistent with the state's EPA-approved water quality standards. Hence, when making an determination based on comparison of ambient data and other information to a numeric WQC expressed as an "average" concentration over a specified period of time, a statement of a desired number of samples may be appropriate. Still, the methodology should provide decision rules for concluding nonattainment in cases where the target data quantity expectations are not met, but the available data and information indicate a reasonable likelihood of a WQC exceedance (e.g., available samples with major digressions from the criterion concentration, corroborating evidence from independent lines of evidence such as biosurveys or incidence of waterborne disease, indications that conditions in the waterbody and loadings of the pollutant into the waterbody have remained fairly stable over the period in question).

Even a very small set of samples may be sufficient to indicate impairment, particularly when the duration/averaging periods of relevant WQC are quite short (an hour or less). For example, one grab sample meeting QA/QC specifications with a concentration higher than the criterion - concentration for a toxic compound could well be grounds for concluding that a WQC expressed as a concentration not be surpassed at any time had been exceeded. A single sample with a concentration that digressed from (was above) the criterion-concentration would be a particularly strong indicator of exceedance of such a criterion if it was the only sample that had been collected. In such a situation, the rate of digression in the sample set (in this instance a set of one) was 100%. This means that, if the timing of the sample was picked randomly, the chances are good that if additional samples had been taken over the period of concern, the vast majority of those would also have had concentrations above the criterion-concentration. (Of course, if the sole sample were collected during a time, condition and/or location condition excluded from application of said WQC, by the state's WQS regulation, it would not be an appropriate basis for 303(d)-listing a segment. Commonly encountered examples of such exclusions include streamflows below the low-flow 7Q10 or areas inside the designated mixing zone for an NPDES permittee.) *NOTE*: See Sec. IV.D.2 for discussion of a somewhat different issue regarding use of single grab samples.

4. Providing Excluded Data (considered and evaluated, but not used) to EPA

EPA regulations require states to provide as part of their section 303(d) list submission a rationale for not using any existing and readily available water quality-related data and information in developing the list. 40 CFR 130.7(b)(6)(iii). EPA recommends that states provide such a rationale on a segment-specific basis to assist EPA in reviewing the state's listing decisions. EPA may also request that states provide any data or information they decided not to use to develop their list and a case-specific rationale for that decision to not use the data in a particular WQS attainment status determination. EPA may review the data and rationale, disapprove section 303(d) listing decisions if appropriate, and make changes in the section 303(d) list based on data and information that was improperly excluded. Failure by a state to provide a reasonable technical rationale for a specific determination or for a decision not to use particular data or information may result in partial disapproval of the list for failure to include segments in Category 5, and potential additions of segments to the section 303(d) list by EPA.

E. Should a state use information other than site-specific ambient monitoring data?

Yes, as appropriate. Categorization decisions should generally not be based only on site specific ambient monitored data, and what was directly observed in the limited set of samples available to the water quality assessor, when other relevant types of information are available. For example, EPA regulations require that "reports from dilution calculations and predictive modeling" be included in the data and information that a state considers in its assessment process for section 303(d) listing (Category 5) purposes (40 CFR 130.7(b)(5)(ii)). Likewise, it may be appropriate to place a segment in any of the other four IR categories based on assessments resulting from the consideration of assessment tools such as predictive modeling, remote sensing data, land use analysis, knowledge about pollutant sources and loadings, observed effects, etc. (see longer listing of types of data and information in Sec. IV, Part C).

EPA believes that a valid assessment of a segment's condition should involve drawing conclusions beyond those which would be arrived at by taking into account nothing more that what was directly observed in the fraction of all possible segment conditions over a given span of time and volume of space represented by a typical set of ambient data. Simple dilution calculations, for example, can be used to estimate what concentration of a pollutant might be present under conditions (e.g., streamflow, pollutant loads) different from those extant at the times sampling was performed.

F. How should states use results of probability-based monitoring?

States should report the results of probability-based assessments as a component of their Integrated Report. A probability-based monitoring design is a type of sample survey design that ensures monitoring at a representative set of sample sites from which inferences can be made about the larger population or resource under investigation (e.g., rivers and streams throughout a state or watershed). It is similar to an opinion poll in which a sample of people are selected at random to represent a larger population. Probability-based designs are used in a wide range of disciplines when conducting a census (e.g., sampling every stream mile) is not economically feasible or is not necessary.

States are encouraged to use probability-based monitoring designs for developing probabilistic statements about waterbody conditions over broad scales (basins, the entire state). EPA believes that a probability-based monitoring design applied over large areas, such as an entire state or a large watershed, is a cost-effective approach to producing a statistical statement, of known confidence, describing the aggregate condition of water resources. For instance, based on such a study, a state might be able to state, with 75% confidence, that 37% of lakes of 50 acres or less fail to meet WQC for total phosphorus.

In addition, sampling performed under probability-based surveys provides site-specific data about each sample location. These data should be considered along with any other site-specific data that might be available, to determine if they should be used to make WQS attainment status determinations, leading to placing segments in the five categories.

The results of probability-based monitoring provide a useful benchmark for the extent that segments are likely to be healthy or degraded. This may help states refine their understanding of how much additional targeted monitoring is needed to complete identification of segments needing restoration, as well as high quality waters needing extra protection.

Though probability-based monitoring programs usually result only in a statistical statement, of known confidence, about aggregate waterbody conditions across a large area, in some instances, results may be compelling enough to support site-specific decisions about water quality in segments besides those from which ambient data were collected. For example, if a probability-based survey of fish tissue from a random sample of lakes across a state found, with a reasonably high level of confidence, that a very high percentage of lakes contain fish with tissue contaminant concentrations exceeding advisory levels, decision makers might decide to list all of the state's lakes as impaired for fish consumption use However, see Section V.H.6 below for additional guidance on use of fish consumption advisories in attainment determinations. For more information on the design and implementation of probability-based sample surveys, visit EPA's Aquatic Resource Monitoring web page at http://www.epa.gov/nheerl/arm/index.htm.

G. How should statistical approaches be used in attainment determinations?

The state's methodology should provide a rationale for any statistical interpretation of data for the purpose of making an assessment determination.

1. Description of statistical methods to be employed in various circumstances

The methodology should provide a clear explanation of which analytic tools the state uses and under which circumstances. EPA recommends that the methodology explain issues such as the selection of key sample statistics (arithmetic mean concentration, median concentration, or a percentile), null and alternative hypotheses, confidence intervals, and Type I and Type II error thresholds. The choice of a statistic tool should be based on the known or expected distribution of the concentration of the pollutant in the segment (e.g., normal or log normal) in both time and space.

Past EPA guidance (1997 305(b) and 2000 CALM) recommended making non attainment decisions, for "conventional pollutants" — TSS, pH, BOD, fecal coliform bacteria, and oil and grease[13] — when more than "10% of measurements exceed the water quality criterion." (However, EPA guidance has not encouraged use of the "10% rule" with other pollutants, including toxics.) Use of this rule when addressing conventional pollutants, is appropriate if its application is consistent with the manner in which applicable WQC are expressed. An example of a WQC for which an assessment based on the ten percent rule would be appropriate is the EPA acute WQC for fecal coliform bacteria, applicable to protection of water contact recreational use. This 1976-issued WQC was expressed as, "...no more than ten percent of the samples exceeding 400 CFU per 100 ml, during a 30-day period." Here, the assessment methodology is clearly reflective of the WQC.

[13] There are a variety of definitions for the term "conventional pollutants." Wherever this term is referred to in this guidance, it means "a pollutant other than a toxic pollutant."

On the other hand, use of the ten percent rule for interpreting water quality data is usually not consistent with WQC expressed either as: 1) instantaneous maxima not to be surpassed at any time, or 2) average concentrations over specified times. In the case of "instantaneous maxima (or minima) never to occur" criteria use of the ten percent rule typically leads to the belief that segment conditions are equal or better than specified by the WQC, when they in fact are considerably worse. (That is, pollutant concentrations are above the criterion-concentration a far greater proportion of the time than specified by the WQC.) Conversely, use of this decision rule in concert with WQC expressed as average concentrations over specific times can lead to concluding that segment conditions are worse than WQC, when in fact they are not.

If the state applies different decision rules for different types of pollutants (e.g., toxic, conventional, and non-conventional pollutants) and types of standards (e.g., acute vs. chronic criteria for aquatic life or human health), the state should provide a reasonable rationale supporting the choice of a particular statistical approach to each of its different sets of pollutants and types of standards.

2. Elucidation of policy choices embedded in selection of particular statistical approaches and use of certain assumptions

EPA strongly encourages states to highlight policy decisions implicit in the statistical analysis that they have chosen to employ in various circumstances. For example, if hypothesis testing is used, the state should make its decision-making rules transparent by explaining why it chose either "meeting WQS" or "not meeting WQS" as the null hypothesis (rebuttable presumption) as a general rule for all waters, a category of waters, or an individual segment. Starting with the assumption that a water is "healthy" when employing hypothesis testing means that a segment will be identified as impaired, and placed in Category 4 or 5, only if substantial amounts of credible evidence exist to refute that presumption. By contrast, making the null hypothesis "WQS not being met" shifts the burden of proof to those who believe the segment is, in fact, meeting WQS.

Which "null hypothesis" a state selects could likely create contrasting incentives regarding support for additional ambient monitoring among different stakeholders. If the null hypothesis is "meeting standards," there were no previous data on the segment, and no additional existing and readily available data and information are collected, then the "null hypothesis" cannot be rejected, and the segment would not be placed in Category 4 or 5. In this situation, those concerned about possible adverse consequences of having a segment declared "impaired" might have little interest in collection of additional ambient data. Meanwhile, users of the segment would likely want to have the segment monitored, so they can be ensured that it is indeed capable of supporting the uses of concern. On the other hand, if the null hypothesis is changed to "segment not meeting WQS," then those that would prefer that a particular segment not be labeled "impaired" would probably want more data collected, in hopes of proving that the null hypothesis is not true.

Another key policy issue in hypothesis testing is what significance level to use in deciding whether to reject the null hypothesis. Picking a high level of significance for rejecting the null hypothesis means that great emphasis is being placed on avoiding a Type I error (rejecting the null hypothesis, when in fact, the null hypothesis is true). This means that if a 0.10 significance level is chosen, the state wants to keep the chance of making a Type I error at or below ten percent. Hence, if the chosen null hypothesis

is "segment meeting WQS," the state is trying to keep the chance of saying a segment is impaired – when in reality it is not – under ten percent.

An additional policy issue is the Type II errors (not rejecting the null hypothesis, when it should have been). The probability of Type II errors depends on several factors. One key factor is the number of samples available. With a fixed number of samples, as the probability of Type I error decreases, the probability of a Type II error increases. States would ideally collect enough samples so the chances of making Type I and Type II errors are simultaneously small. Unfortunately, resources needed to collect such numbers of samples are quite often not available.

The final example of a policy issue that a state should describe is the rationale for concentrating limited resources to support data collection and statistical analysis in segments where there are documented water quality problems or where the combination of nonpoint source loadings and point source discharges would indicate a strong potential for a water quality problem to exist.

EPA recommends that, when picking the decision rules and statistical methods to be utilized when interpreting data and information, states attempt to minimize the chances of making either of the two following errors:

- Concluding the segment is impaired, when in fact it is not, and
- Deciding not to declare a segment impaired, when it is in fact impaired.

States should specify in their methodology what significance level they have chosen to use, in various circumstances. The methodology would best describe in "plain English" the likelihood of deciding to list a segment that in reality is not impaired (Type I error if the null hypothesis is "segment not impaired"). Also, EPA encourages states to estimate, in their assessment databases, the probability of making a Type II error (not putting on the 303(d) list a segment that in fact fails to meet WQS), when: 1) commonly-available numbers of grab samples are available, and 2) the degree of variance in pollutant concentrations are at commonly encountered levels. For example, if an assessment is being performed with a WQC expressed as a 30-day average concentration of a certain pollutant, it would be useful to estimate the probability of a Type II error when the number of available samples over a 30 day period is equal to the average number of samples for that pollutant in segments state-wide, or in a given group of segments, assuming a degree of variance in levels of the pollutant often observed over typical 30 day periods.

H. How should states use community-level bioassessment data?

Many states use multi-metric, community-level biological assessments to report water resource condition. Biological assessments provide direct measures of the cumulative response of the biological community to all sources of stress. Therefore, a biocriteria/bioassessment represents a very useful indicator of the use support status for aquatic life.

Credible assessments of biological condition can be accomplished with far fewer samples than with parameter-specific monitoring. However, attention to proper quality assurance and control is equally important in biological monitoring as it is in chemical and physical measurements. Threshold values for segment impairment determinations as well as quality assurance should be addressed in the state's methodology.

States should include biological assessments in the data and information they assemble and evaluate in developing their Integrated Reports, and must provide a rationale for any decision not to use the assessments in developing their section 303(d) lists.

States using biological assessments to make reporting determinations should also consider other types of data and information (i.e., chemical and physical). In instances in which the indication of aquatic life use support provided by biosurvey data and that provided by chemical and/or physical data differ, EPA continues to support the principle of independent applicability (see Section IV.K. below), as most recently articulated in its Consolidated Assessment and Listing Methodology guidance.

I. What information should the state provide regarding its interpretation of its WQS?

When deciding whether to put a segment in Category 1, 2, 4 or 5, a state is trying to answer the question, "What does available ambient monitoring data and other information tell us about whether or not this segment is meeting WQS?" In order to answer this question, it is necessary to be very clear about the WQS that apply to the segment– the DUs assigned to the segment, as well as the numeric water quality criteria (WQC) applicable to each DU, along with narrative WQC.

Ideally, states' WQS regulations will clearly articulate each DU and all WQC applicable to that DU. However, in some instances, there may be ambiguity in the way WQS are expressed. For example, a WQC could refer to an "average" concentration. This could mean the median, the arithmetic mean, the geometric mean, or something else describing a central tendency. Also, WQS regulations and guidance sometimes do not clearly state a duration component of a WQC (criterion-duration) – particularly some types of human health (HH) criteria. (For reference purposes: EPA HH criteria for carcinogens are presumed to have a duration of a year or more; whereas a duration of 30 days is employed in criteria addressing human pathogens and water contact recreation. EPA's aquatic life WQC for toxic chemicals present acute and chronic concentrations applicable to exposure durations of (a) 1 hour and (b) 4 days, respectively.)

For toxic ("priority" pollutants) and protection of freshwater aquatic life, EPA guidance recommends use of a once in three year maximum allowable excursion recurrence frequency (*Guidance for 2004 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d) and 305(b) of the Clean Water Act,* Section III (F), EPA, 2003). Hence, for example, if a state freshwater aquatic life WQC were expressed as "1-hour average concentration not to surpass 22 ug/L", EPA would assume an applicable excursion frequency of no more than once in 3 years.

J. How should states handle shared waters?

States with shared waters should make every effort to coordinate with each other in the development of their Integrated Reports. Coordination should occur early in the process. Where possible, states should work together to collect, assemble, solicit, and assess all readily available data and information relevant to the shared waters. Assessments for waters that are shared by neighboring states should be as consistent as possible. This is particularly important for segments listed in Category 5. However, differing state WQS can make consistent attainment decisions difficult. In such cases, EPA

Regional offices and interstate commissions, where applicable, should assist in resolving inconsistencies when they arise. The Integrated Report should document the coordination that has occurred between neighboring states and interstate commissions.

Some interstate commissions are required to prepare a section 305(b) report, but the responsibility of preparing Integrated Reports and section 303(d) lists rests with the states. Data and information in an interstate commission section 305(b) report should be considered by the states as one source of readily available data and information when they prepare their IR and make decisions on segments to be placed in Category 5; however, data in a section 305(b) interstate commission report should not be automatically entered in a state IR or section 303(d) list without consideration by the state about whether such inclusion is appropriate. EPA has made the necessary modification to its ADB system to ensure that interstate commission data stays segregated from state data.

K. How does the state make attainment decisions when different types of data indicate a different attainment status?[14]

To address the possibility of conflicting results among different types of data used to assess attainment with WQS, EPA recommends that states apply the policy on independent applicability as appropriate for making WQS attainment decisions. This policy was initially crafted to address development of NPDES permit discharge limits. Its use is slightly different in the context of WQS attainment decisions.

The intent of this policy is to protect against dismissing valuable information when evaluating aquatic life use support, particularly in detecting impairment. EPA's policy on independent application is based on the premise that any valid, representative dataset indicating an actual or projected water quality impairment should not be ignored when one is determining the appropriate action to be taken. However, EPA recognizes that there are circumstances when conflicting results should be investigated further before the attainment or nonattainment decision is made. For example, states may obtain multiple datasets of varying quality, which may influence the reliability of the assessment results.

Figure 4-1 elaborates on the use of the independent application policy in reconciling conflicting results among different datasets used to assess attainment with aquatic life-based WQS. The decision process begins in the upper left of Figure 4-1. When a state, territory, or authorized tribe has two or more types of data that do not indicate consistent attainment status, it should determine whether differences in assessment results can be attributed to differences in the quality of the datasets. For example, this may involve consideration of analytical methods, review of sampling techniques, and detailed assessment of datasets. When the differences are due to data quality issues, the independent application policy allows for resolving the differences by cleaning the data or weighing the higher quality dataset more favorably in the attainment decision.

[14]Part K of Section IV has been adapted from pages 3-9 to 3-10 of EPA's *Consolidated Assessment and Listing Methodologies (CALM)*, July 2002.

When detailed data analysis fails to identify data quality issues that explain the discrepancies, site-specific environmental conditions should be considered (e.g., effects of water chemistry, or the ability of species to adapt over time). Site specific WQC may be explored via application of the water effects ratio, resident species recalculations, or other appropriate methods.

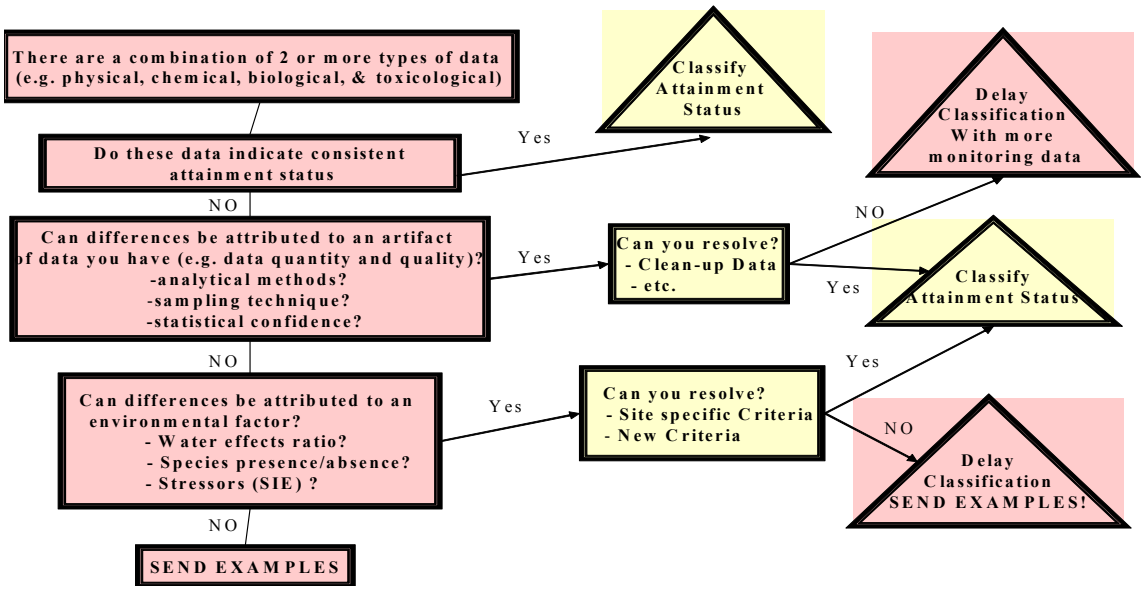

Figure 4-1. Using Multiple Types of Data to Assess Attainment

For Purposes of WQS Attainment/Nonattainment Determinations:

Policy of independent applicability *says*:

- When evaluating multiple types of data (e.g., biological, chemical) and any one type of data indicates an element of a WQS is not attained, the segment should most likely be identified as impaired.

- If there is reason to doubt the nonattainment finding, re-evaluate all of the data sets to resolve discrepancies. In some cases this may lead to modification of applicable WQS to account for site-specific information.

Policy of independent applicability *does not say*:

- Always assume that a single sample result showing impairment outweighs all other data showing attainment.

- Accept all differences in data findings at face value.